PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	**FOR FURTHER ACTION** see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
A25830 WO		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/GB 00/ 04206	02/11/2000	02/11/1999
Applicant		
BRITISH TELECOMMUNICATIONS public limited company		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 3 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language,** the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

   [ ] the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

   b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

   [ ] contained in the international application in written form.

   [ ] filed together with the international application in computer readable form.

   [ ] furnished subsequently to this Authority in written form.

   [ ] furnished subsequently to this Authority in computer readble form.

   [ ] the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

   [ ] the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. [ ] **Certain claims were found unsearchable** (See Box I).

3. [ ] **Unity of invention is lacking** (see Box II).

4. With regard to the **title,**

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract,**

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. 5

   [X] as suggested by the applicant.

   [ ] because the applicant failed to suggest a figure.

   [ ] because this figure better characterizes the invention.

   [ ] None of the figures.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 May 2001 (10.05.2001)

PCT

(10) International Publication Number
WO 01/33554 A1

(51) International Patent Classification[7]: G10L 15/26

(21) International Application Number: PCT/GB00/04206

(22) International Filing Date:
2 November 2000 (02.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
99308680.0 2 November 1999 (02.11.1999) EP

(71) Applicant *(for all designated States except US)*: **BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY** [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **MILNER, Benjamin, Peter** [GB/GB]; 9 The Fairway, Gorleston-on-Sea, Great Yarmouth, Norfolk NR31 6JS (GB).

(74) Agent: **SEMOS, Robert, Ernest, Vickers**; BT Group Legal Services, Intellectual Property Dept., 120 Holborn, Holborn Centre, 8th floor, London EC1N 2TE (GB).

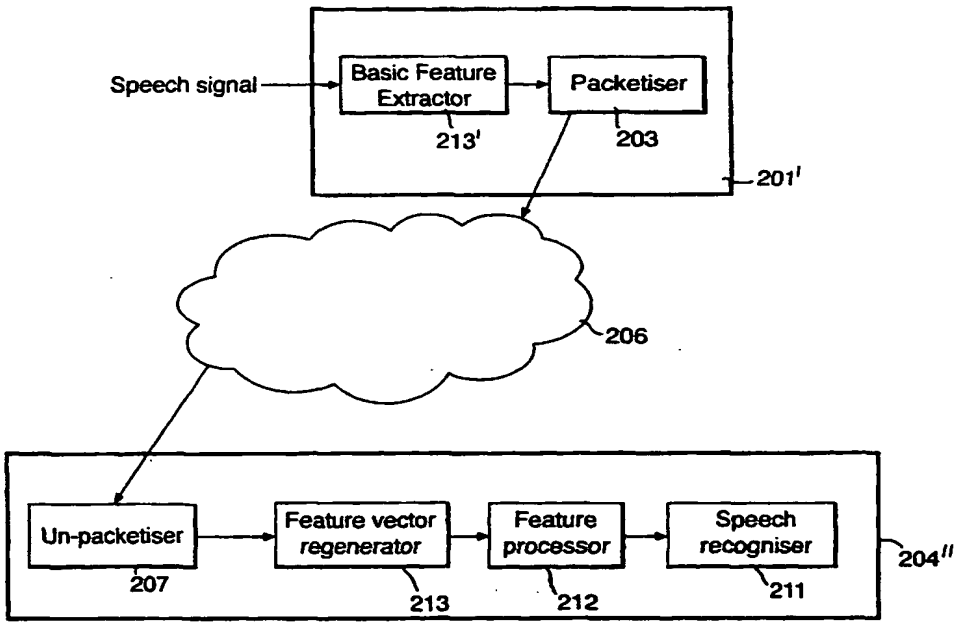
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SPEECH RECOGNITION



(57) Abstract: A speech recogniser suitable for distributed speech recognition is robust to missing speech feature vectors. Speech is transmitted via a packet switched network in the form of basic feature vectors. Missing feature vectors are detected and replacement feature vectors are estimated by interpolation of received data prior to speech recognition. In an improved version features are converted and interpolation is done in a spectral domain.

WO 01/33554 A1

Rec'd PCT/PTO 02 APR 2002
PCT/GB00/04206
10/089562

SPEECH RECOGNITION

This invention relates to a method of and an apparatus for speech recognition which is robust to missing speech data. It is particularly useful in distributed speech recognition in which data is transmitted via a packet switched network.

Recently there has been an enormous increase in the use of mobile devices such as mobile phones and personal digital assistants. It is desirable to make the human to device interface as natural and easy to use as possible. Speech recognition is one solution which increases naturalness, and overcomes the difficulties in using very small keyboards found on many mobile devices. A Personal Computer (PC) usually provides sufficient processing power to operate a speech recogniser. However, on mobile devices processing power is a limiting factor. One solution is to use distributed speech recognition (DSR). DSR makes use of remote speech recognisers which are accessed by a device across a transmission network. Speech data from the device is transmitted across the network to the remote speech recogniser and the remote speech recogniser processes the speech to provide a recognition result (or set of results) which is then transmitted back to the device.

There are basically two types of network across which such information can be transmitted; namely connection-orientated networks and connectionless networks. The connection-orientated network is essentially the telephony service which has evolved over the last 100 years for the switching and transmission of voice data. A connectionless network is packet-based and its main functionality is the routing and switching of data packets from one location to another.

When a call is made on a connection-orientated network a reservation is made to ensure that sufficient network resources are available to sustain the call. This may be the allocation of a physical connection or of time slots in a pulse code modulation (PCM) system. If sufficient resources are not available then the call is refused, typically accompanied by an engaged signal.

The connectionless network is very much aimed at the routing and switching of data packets and is designed to efficiently handle the high burstiness of this traffic. Packets are comprised of two parts – a header and payload. The header contains information regarding the source and destination address while the payload contains the actual data which needs to be sent.

For transmitting real-time data such as speech, the essential difference between the two networks is that the connection-orientated network reserves sufficient capacity, or bandwidth, to maintain a connection throughout the call. With a connectionless network sufficient bandwidth is not guaranteed which means that the network may produce delays or missing packets which interrupt the data transmission. Therefore the connection-orientated network is much better suited to delivering real-time data. Voice has therefore traditionally been transmitted using connection-orientated networks. However, because of the enormous growth in data networks, the technique of Voice over Internet Protocol (VoIP) has been developed to allow the real-time transmission of voice signals across connectionless networks.

In a connectionless network the packets containing the speech can be routed across a wide variety of paths depending on the network traffic. Indeed, it may be that successive packets are routed around the network on different paths. As a result it is possible that some packets arrive out of sequence or may never even arrive. This is clearly undesirable in a DSR system as it will introduce recognition errors. An approach to dealing with this problem of missing packets is to use protocols designed specifically for real-time data which ensure all the data arrives with minimal delay.

The traditional connectionless network is termed best-effort. This means that packets from a source are sent to a destination with no guarantee of a timely delivery. For applications such as file transfer which require a guarantee of delivery, Transmission Control Protocol (TCP) is able to trade packet delay for guaranteed reception. In the event of lost packets TCP allows for the destination to request the retransmission of those lost packets. However, for real-time data it is important to minimise transmission delays. It is therefore impracticable to use TCP and wait for the retransmission of lost packets. A better approach is to use User Datagram Protocol (UDP) as the protocol for sending the packets. This has a short duration buffer which allows for slight delays in packet arrival after which UDP assumes the packet is not going to arrive. No facility for the retransmission of lost packets is available. This has the advantage that delays are minimised but at the expense of possibly losing some of the speech signal when network traffic is high and packet loss is probable.

Protocols designed specifically for real-time data transmission include Resource Reservation Protocol (RSVP). This is a signalling protocol which reserves network resources at the start of a call to ensure that a direct connection to the destination is available throughout. In effect it makes a connection-orientated path from

a connectionless network. In order for this to function all the routers in the network from the source to destination must be RSVP enabled. As RSVP is a relatively new protocol not all routers are equipped with this facility.

Another protocol designed specifically for real-time data transmission is DiffServ. This makes use of a byte of data in the packet header to specify a Type of Service (ToS) – i.e. how much priority should be given to the immediate routing of that packet through the network. Clearly some data will have very high priority such as network management and system commands. Lower priority will be given to file transfer and email where immediate delivery is not too important. Depending on the emphasis given to the network, high priority can be given to speech packets to assist real-time use. Again, this protocol is only in development and not available generally.

The increase of connectionless voice networks, coupled with the increase in automation of call centres means that the ability to perform robust speech recognition over a connectionless network is becoming more important.

An alternative approach to ensuring that all packets containing the speech signal successfully reach the speech recogniser is to make the recogniser itself robust to missing packets. When the packet loss is low (<5%) the drop in recognition performance is not too significant. However, as packet loss increases – or occurs in bursts – the effect is more detrimental. Therefore, a speech recogniser is required which is able to tolerate this loss of speech.

Known signal processing techniques which deal with missing packets range from very simple to complex – a good review is made in C. Perkins, O. Hodson and V. Hardman, "A survey of packet loss recovery techniques for streaming audio", IEEE Network Magazine, Vol.12, No. 5, pp. 40-48, October 1998. Simple techniques include splicing which merely joins the speech signal together either side of the gap. Silence and noise substitution replace the missing frames of speech with either silence or noise. Repetition replaces the lost frames of speech with copies of the speech which arrived before the gap.

More sophisticated techniques attempt to estimate the missing parts of the signal from those parts which have been correctly received. These include waveform substitution which uses the pitch on either side of the gap to estimate the missing speech. Time scale modification stretches the audio signal either side of the gap to fill in the missing speech. Regeneration-based repair uses parameters of the codec to

determine the required fill-in speech. All these techniques attempt to reconstruct the time-domain speech signal.

According to the present invention there is provided a method of speech recognition comprising the steps of receiving a sequence of transmitted feature vectors, said feature vectors representing a speech signal; detecting the absence of a feature vector in the received sequence; generating an estimated replacement feature vector for the detected absent feature vector; inserting said replacement feature vector into the received feature vector sequence to provide a modified feature vector sequence; and performing speech recognition upon the modified feature vector sequence.

Preferably the feature vector comprises a plurality of components and the generating step comprises estimating a component of a replacement feature vector by interpolating the corresponding component of a received feature vector.

In a preferred embodiment the estimating step uses an interpolation coefficient corresponding to a component of the received feature vector and further comprising the step of updating the interpolation coefficient in response according to one or more received feature vectors.

In an alternative embodiment of the invention the transmitted feature vectors include features in a cepstral domain, and in which the estimating step comprises the sub steps of converting a received feature vector to a spectral domain; estimating a spectral component by interpolating the corresponding component of the converted feature vector; and converting the estimated spectral component to said cepstral domain.

According to another aspect of the invention there is provided a device for performing speech recognition upon a sequence of parameterised feature vectors comprising a missing feature vector detector arranged in operation to receive the transmitted feature vectors and to indicate the absence of a feature vector in the received sequence; a feature vector estimator arranged, in operation, to receive transmitted feature vectors and responsive to said indication from the missing feature vector detector to estimate a replacement feature vector; a sequence reconstructer arranged, in operation, to receive transmitted feature vectors and to receive a replacement feature vector and to provide as an output a modified feature vector sequence; and a speech recogniser arranged, in operation, to receive the modified feature vector sequence.

Preferably the feature vector estimator comprises an interpolator arranged to receive a feature vector and to provide as an output a component of the replacement feature vector.

In a preferred embodiment in the interpolator uses an interpolation coefficient corresponding to a component of the received feature vector and in which the interpolator is arranged to update the interpolation coefficient in response to receipt of a feature vector.

In an alternative embodiment of the invention the feature vector estimator comprises a first converter for converting a received feature vector to a spectral domain; an estimator for estimating a spectral component by interpolating the corresponding component of the converted frame; a second converter for converting the estimated spectral component to said cepstral domain.

A data carrier loadable into a computer and carrying instructions for causing the computer to carry out a method according to the invention and a data carrier loadable into a computer and carrying instructions for enabling the computer to provide the device according to the invention are also provided.

Embodiments of the invention will now be described, by way of example only, with reference to the accompanying drawings in which:

Figure 1 is a schematic representation of a computer loaded with software embodying the present invention;

Figure 2 is a functional block diagram showing program elements for software embodying a known technique for performing DSR;

Figure 3 is a functional block diagram showing program elements for software embodying another known technique for performing DSR;

Figure 4 is a functional block diagram of program elements that comprise a parameteriser of Figures 2 and 3;

Figure 5 is a functional block diagram of the program elements that comprise the software indicated in Figure 1;

Figure 6 is a functional block diagram of the program elements which comprise a feature vector regenerator of Figure 5;

Figure 7 is a functional block diagram of the program elements that comprise a frame estimator shown in Figure 6 in one embodiment of the invention; and

Figure 8 is a functional block diagram of the program elements that comprise the frame estimator shown in Figure 6 in a second embodiment of the invention.

Figure 1 illustrates a conventional computer 101, such as a PC, running a conventional operating system 103, such as Windows (a Registered Trade Mark of Microsoft Corporation), and having a number of resident application programs 105 including a word processing program, a network browser and e-mail program and a database management program. The computer 101 is also connected to a conventional disc storage unit 111 for storing data and programs, a keyboard 113 and mouse 115 for allowing user input and a printer 117 and display unit 119 for providing output from the computer 101. The computer 101 also has access to external networks (not shown) via a network card 121. The computer 101 also includes a speech recognition program 109 that enables a speech signal received via the network card 121 to be recognised.

In Figure 2, a mobile device 201 includes a framer 205 which divides a received speech signal into short duration frames, for example 30ms, and sends the resultant frames to an encoder 202. The encoder 202 encodes each frame of received speech into a suitable coded representation, for example using the standard codec defined in ITU-G.723.1, and the resultant coded representation is sent to a packetiser 203. The coded representation forms the payload for a packet (not shown) which has a header added by the packetiser 203. The packet is transmitted via a connectionless network 206 to a remote device 204. The remote device 204 includes an unpacketiser 207 which removes the header, and a decoder 208 which decodes the coded representation of the speech frame. Speech frames are sent from the decoder 208 to an audio reconstructor 209 where the speech signal is reconstructed. The speech signal is then parameterised by a parameteriser 210 to form feature vectors suitable for use by a speech recogniser 211. The parameteriser 210 comprises a basic feature extractor 213 and a feature processor 212, the operation of which will be described later.

Figure 3 shows a system for DSR which avoids encoding and decoding the speech signal of the speech by transmitting parameterised speech signals over the network 206. In Figure 3 a device 201' includes a basic feature extractor 213' which parameterises speech signals to form feature vectors. The speech feature vectors are packetised by the packetiser 203 and transmitted via the network 206 to a remote device 204'. At the remote device the features are un-packetised by the unpacketiser

207 and transmitted to the speech recogniser 211 via the feature processor 212. This approach is advantageous over the system shown in Figure 2. Encoding and decoding the signal causes a degradation in quality of the speech signal. This causes a significant reduction in speech recognition performance. By parameterising the speech signal before transmission across the network there is no resultant loss in speech recognition performance. As encoding and decoding of the speech signal is not required there is a significant saving in computation.

The problem of missing packets needs to be addressed. In this invention reconstructing the transmitted feature vector sequence is performed by detecting mising feature vectors and subsequently estimating corresponding replacement feature vectors.

Figure 5 shows a remote device 204" according to the invention, which is implemented on a conventional computer as illustrated in Figure 1. After received features have been unpacketised by the unacketiser 207, missing features are restored by a feature vector regenerator 214.

In the embodiment of the invention described here the basic feature vectors used are Mel-frequency cepstral coefficients (MFCCs). MFCCs are generated from a received speech signal as illustrated in Figure 4. A high emphasis filter 10, normally referred to as a pre-emphasis filter, receives a digitised speech waveform at, for example, a sampling rate of 8 kHz as a sequence of 8-bit numbers and performs a high emphasis filtering process (for example by executing a $1 - 0.95.z^{-1}$ filter), to increase the amplitude of higher frequencies.

A sequence of 256 contiguous samples (referred to as a frame in this description) of the filtered signal is windowed by a window processor 11 in which the samples are multiplied by predetermined weighting constants using, in this embodiment of the invention, a Hamming window, to reduce spurious artefacts generated at the edges of the frame. Each frame overlaps with neighbouring frames by 50%, so as to provide one frame every 16ms.

Each frame of 256 windowed samples is then processed by a MFCC generator 12 to extract an MFCC feature vector comprising eight MFCC's.

The MFCC feature vector is derived by performing a spectral transform, in this embodiment of the invention, a Fast Fourier Transform (FFT), on each frame of a speech signal, to derive a representation of the signal spectrum for each frame of

speech. The terms of the spectrum are integrated into a series of broad bands, which are distributed in a 'mel-frequency' scale along the frequency axis, to provide nineteen mel-frequency features. These features are referred to as filterbank features in this description. The mel-frequency scale is a perceptually motivated scale, which comprises frequency bands evenly spaced on a linear frequency scale between 0 and 1 kHz, and evenly spaced on a logarithmic frequency scale above 1 kHz. The logarithm of each mel-frequency feature is calculated and then a Discrete Cosine Transform (DCT) is performed to generate an MFCC feature vector for the frame. Features such as the mel-frequency features described above, which represent the frequencies within a signal are referred to as being features in a spectral domain. Features which represent the rate of change of frequencies in a signal, such as the MFCC's described above are referred to as being in a cepstral domain.

For MFCC's it is found that the useful information is generally confined to the lower order coefficients, so in this embodiment of the invention nine cepstral coefficients are used.

Before the features are transmitted to the feature processor 212, any missing front end feature vectors are restored by the feature vector regenerator illustrated in Figure 6.

Estimation of missing feature vectors is a simpler problem than estimation of the original time-domain speech signal. Feature vectors are highly correlated with one another in time, and represent a longer portion of speech than a single digital time-domain sample. In the embodiment of the invention described here 256 time-domain samples are represented by 9 MFCC's. The estimation of 9 MFCC's which are highly correlated with preceding and following MFCC's is much simpler than accurate estimation of 256 samples.

In Figure 6 a stream of MFCCs is shown with one missing feature vector due to the packet having that feature vector as part of it's payload being lost in the network 206. Received MFCCs are first passed into a missing feature vector detector 501 which identifies whether any feature vectors are missing or not. If a missing feature vector is detected a feature vector estimator 502 is used to estimate the missing feature vector. The feature vector sequence is then reconstructed by the sequence reconstructor 503. The resultant reconstructed sequence may then be used for speech recognition is the usual way.

In the embodiment of the invention described here the missing feature vector detector 501 uses feature vector numbering. An additional feature is added to the feature vector by the basic feature extractor 213', the additional feature indicates the position of each feature vector in the feature vector sequence. At the remote device 204" the missing feature vector detector 501 checks the feature vector number of each feature vector received and uses this number to detect whether there are any missing feature vectors, and if so how many. When one or more missing feature vectors are detected a signal is sent to the feature vector estimator 502.

The feature vector estimator 502 uses interpolation to estimate the missing speech features. Each feature is estimated separately and the time series of each is used to form a polynomial which enables missing elements to be estimated. In this embodiment of the invention a simple straight line interpolation is used. A detailed description of interpolation algorithms is provided in S.V. Vaseghi, "Advanced signal processing and digital noise reduction", John-Wiley, 1996

Figure 7 illustrates interpolation of a MFCC feature vector. For each feature in the feature vector a corresponding interpolator 601, 602, .. 609 is established. As each new feature vector arrives the interpolation coefficients for each feature of the feature vector are updated. When a missing feature vector is detected by the missing feature vector detector 501 an estimate of the missing feature vector is made using the interpolators 601, 602, .. 609. The estimate of the missing frame is then inserted into the feature vector sequence by the sequence reconstructor 503.

In another embodiment of the invention the MFCC feature vectors, which are in the cepstral domain are converted back into the spectral domain so that interpolation is performed on features which represent the frequencies in the original signal. Upon detection of a missing feature vector the interpolator produces an estimate of a filterbank feature vector. This is then logged and a DCT applied to transform the estimate into the MFCC domain. This is illustrated in Figure 8 in which a sequence of received MFCC feature vectors 701 has an inverse DCT applied to it at 702, and is exponentiated at 703 (i.e. the inverse of a logarithm is applied) to provide a sequence of filterbank feature vectors. A filterbank interpolator 705 is used to provide a filterbank estimate 706 of a missing feature vector, and the filterbank estimate 706 has a logarithm calculated at 707 and a DCT applied at 708 to provid an MFCC estimate 709.

After the feature vector sequence has been reconstructed by the feature vector regenerator 214, processing of the basic features prior to recognition is performed by the feature processor 212. RASTA filtering is applied by bandpass filtering the time series of feature vectors. A detailed description of RASTA filtering may be found in H. Hermansky and N. Morgan, "RASTA processing of speech", IEEE Trans. Speech and Audio Proc., vol. 2, no. 4, pp. 578-589, October 1994. Any channel distortion is additive in the cepstral domain, so applying a sharp cut-off highpass filter to each of the features, across time, removes any offset and hence suppresses channel distortion. Cepstral Time Matrix (CTM) features are then calculated by taking a DCT across a sequence of seven MFCCs. A detailed description of CTM features may be found in B.P. Milner, "Inclusion of temporal information into features for speech recognition", Proc. ICSLP, pp. 256-259, 1996

It will be appreciated by those skilled in the art that the technique described could be applied to other types of basic speech parameterisation. Cepstral features may be calculated using a Fourier transform as described here, or using linear predictive (LP) analysis. It can be proven that the resultant cepstrum from either of these two routes is identical. In the embodiment of the invention described here the Fourier transform based cepstrum has been modified to include a mel-scale filterbank resulting in MFCC's.

A process similar to the mel-scale filterbank is used in perceptual linear predictive (PLP) analysis where a set of critical-band filters are convolved with the speech spectrum. These modify the spectrum according to perceptual measurements of human hearing and lead to the PLP cepstrum.

It will also be appreciated by those skilled in the art that other feature vector processing techniques could be applied, for example differential features may be calculated, such as 'velocity' and 'acceleration' of the basic features. Cepstral mean normalisation in which the average of each feature is subtracted from each feature respectively, may be used. Linear discriminant analysis (LDA) as described in E.J. Paris and M.J. Carey, "Estimating linear discriminant parameters for continuous density HMMs", Proc. ICSLP, pp. 215-218, 1994 may also be used.

As will be understood by those skilled in the art, the speech recognition program 109 can be contained on various transmission and/or storage mediums such as a floppy disc, CD-ROM, or magnetic tape so that the program can be loaded onto

one or more general purpose computers or could be downloaded over a computer network using a suitable transmission medium.

Unless the context clearly requires otherwise, throughout the description and the claims, the words "comprise", "comprising" and the like are to be construed in an inclusive as opposed to an exclusive or exhaustive sense; that is to say, in the sense of "including, but not limited to".

CLAIMS

1. A method of speech recognition comprising the steps of

receiving a sequence of transmitted feature vectors, said feature vectors representing a speech signal;

detecting the absence of a feature vector in the received sequence;

generating an estimated replacement feature vector for the detected absent feature vector;

inserting said replacement feature vector into the received feature vector sequence to provide a modified feature vector sequence; and

performing speech recognition upon the modified feature vector sequence.

2. A method according to claim 1, in which the feature vector comprises a plurality of components and the generating step comprises

estimating a component of a replacement feature vector by interpolating the corresponding component of a received feature vector.

3. A method according to claim 2, in which the estimating step uses an interpolation coefficient corresponding to a component of the received feature vector and further comprising the step of

updating the interpolation coefficient in response according to one or more received feature vectors.

4. A method according to claim 1, in which the transmitted feature vectors include features in a cepstral domain, and in which the estimating step comprises the sub steps of

converting a received feature vector to a spectral domain;

estimating a spectral component by interpolating the corresponding component of the converted feature vector; and

converting the estimated spectral component to said cepstral domain.

5. A device for performing speech recognition upon a sequence of parameterised feature vectors comprising

a missing feature vector detector arranged in operation to receive the transmitted feature vectors and to indicate the absence of a feature vector in the received sequence;

a feature vector estimator arranged, in operation, to receive transmitted feature vectors and responsive to said indication from the missing feature vector detector to estimate a replacement feature vector;

a sequence reconstructer arranged, in operation, to receive transmitted feature vectors and to receive a replacement feature vector and to provide as an output a modified feature vector sequence; and

a speech recogniser arranged, in operation, to receive the modified feature vector sequence.

6. A device according to claim 5, in which the feature vector estimator comprises

an interpolator arranged to receive a feature vector and to provide as an output a component of the replacement feature vector.

7. A device according to claim 6, in which the interpolator uses an interpolation coefficient corresponding to a component of the received feature vector and in which the interpolator is arranged to update the interpolation coefficient in response to receipt of a feature vector.

8. A device according to claim 6, in which the feature vector estimator comprises

a first converter for converting a received feature vector to a spectral domain;

an estimator for estimating a spectral component by interpolating the corresponding component of the converted frame;

a second converter for converting the estimated spectral component to said cepstral domain.

9. A data carrier loadable into a computer and carrying instructions for causing the computer to carry out the method according to any one of claims 1 to 4.

10. A data carrier loadable into a computer and carrying instructions for enabling the computer to provide the device according to any one of claims 5 to 8.

11. A method of speech recognition substantially as described herein with reference to Figures 5 to 8.

12. A device for recognising speech substantially as described herein with reference to Figures 5 to 8.

Fig.1.

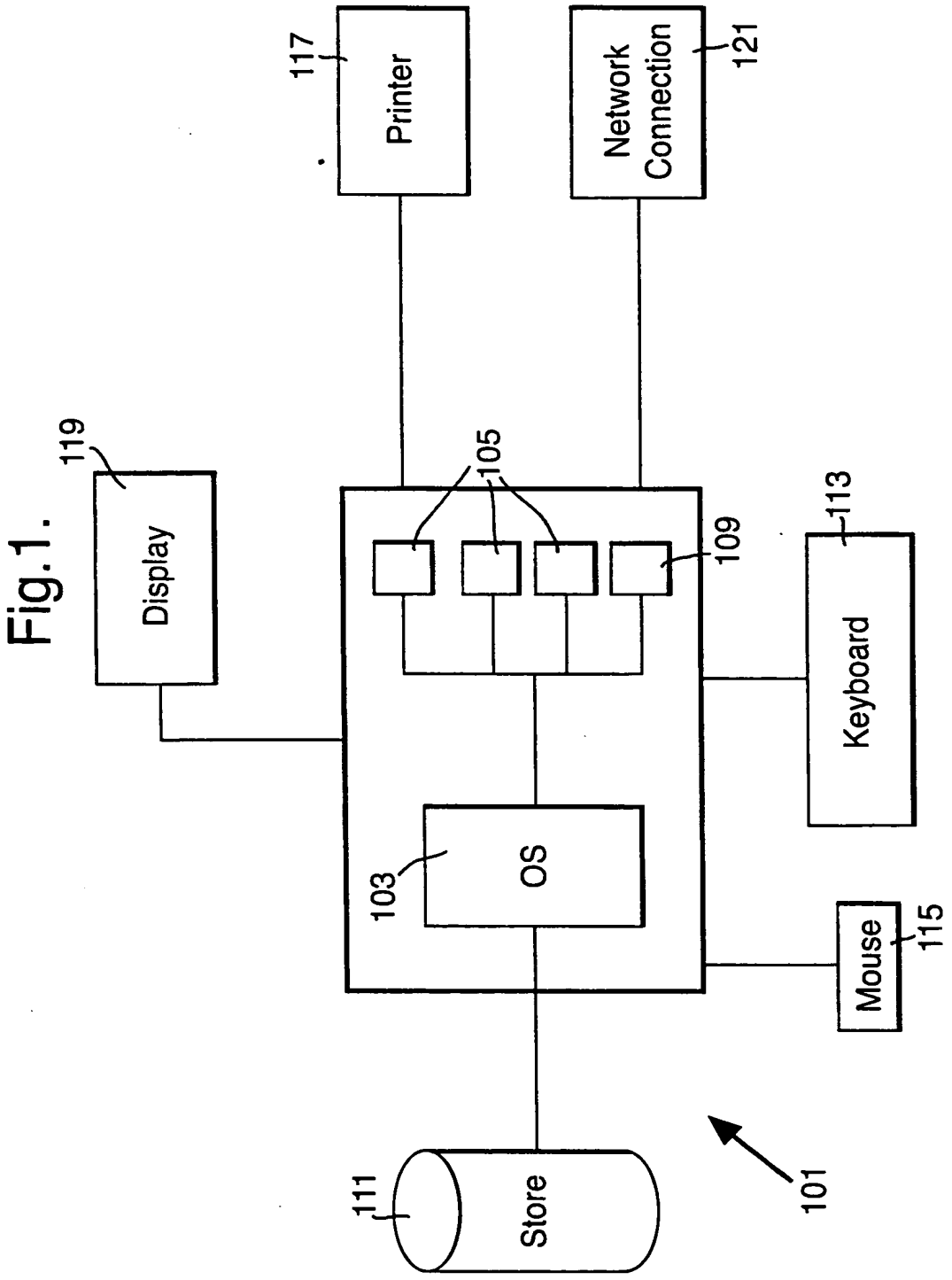

117
Printer
Network Connection
121
119
Display
105
109
113
Keyboard
OS
103
Mouse
115
111
Store
101

Fig.2.

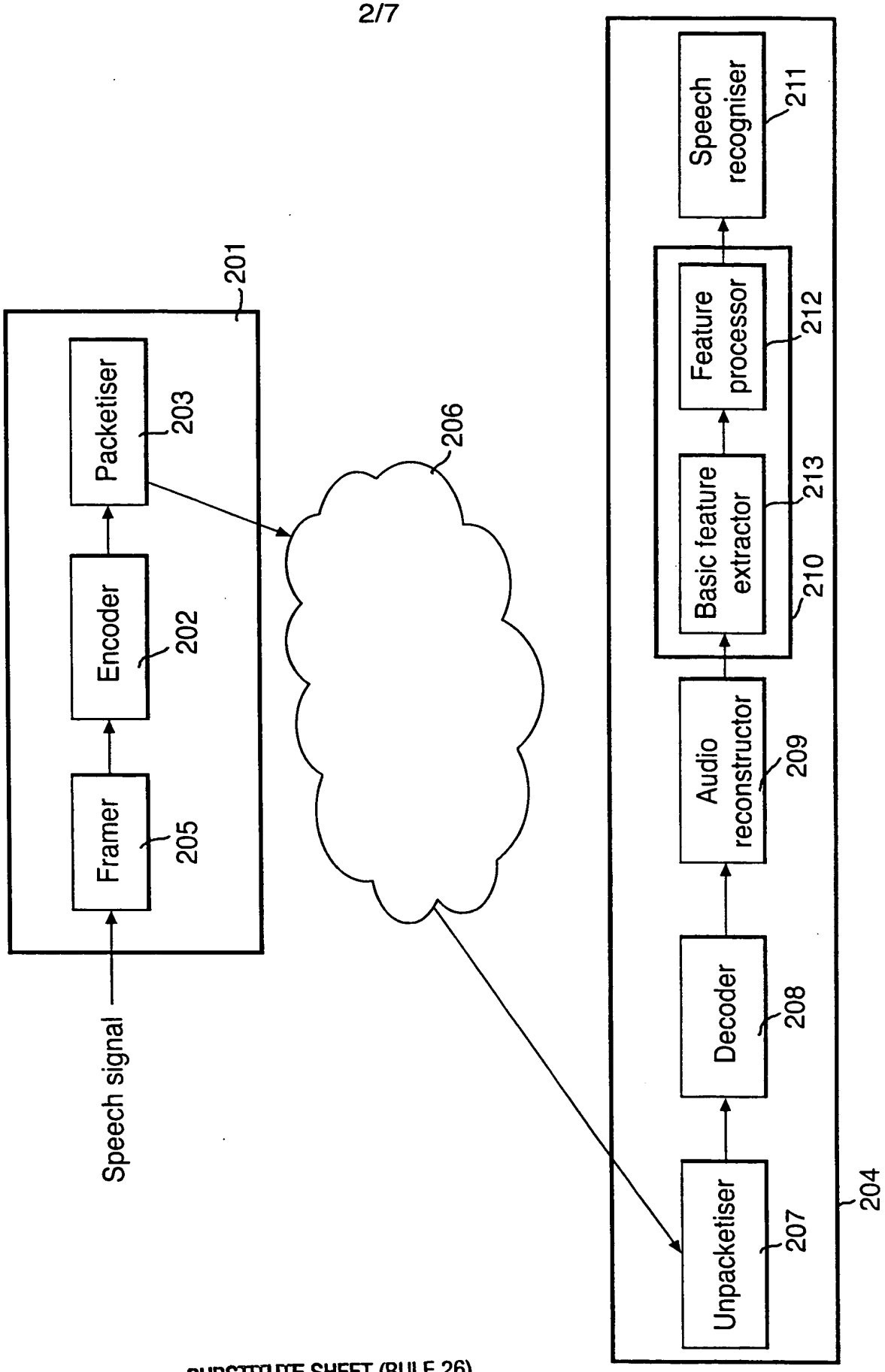

Speech signal
Framer
205
Encoder
202
Packetiser
203
201
206
Unpacketiser
207
Decoder
208
Audio reconstructor
209
Basic feature extractor
213
210
Feature processor
212
Speech recogniser
211
204

Fig.3.

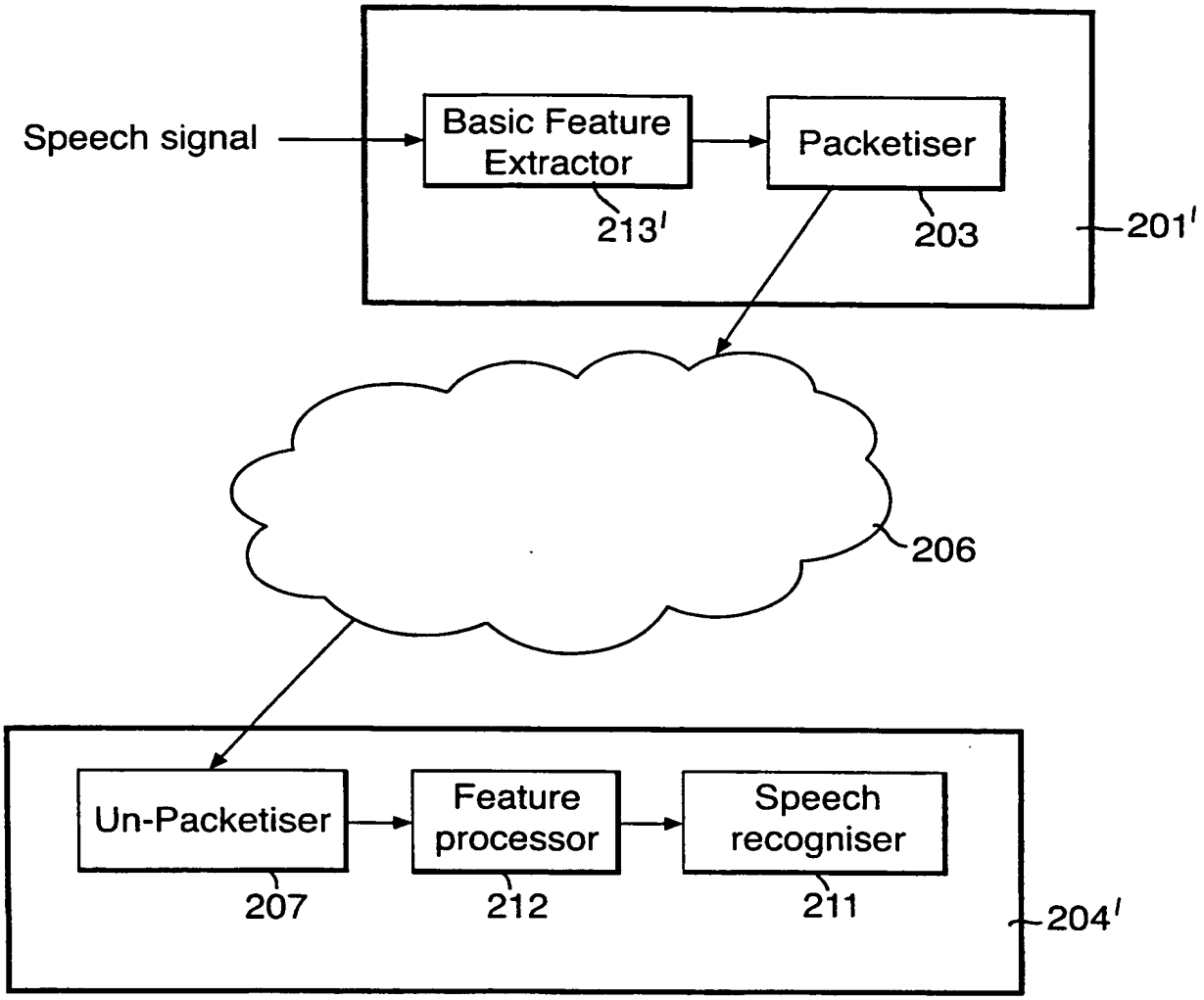

Speech signal
Basic Feature Extractor
213'
Packetiser
203
201'
206
Un-Packetiser
207
Feature processor
212
Speech recogniser
211
204'

Fig.4.

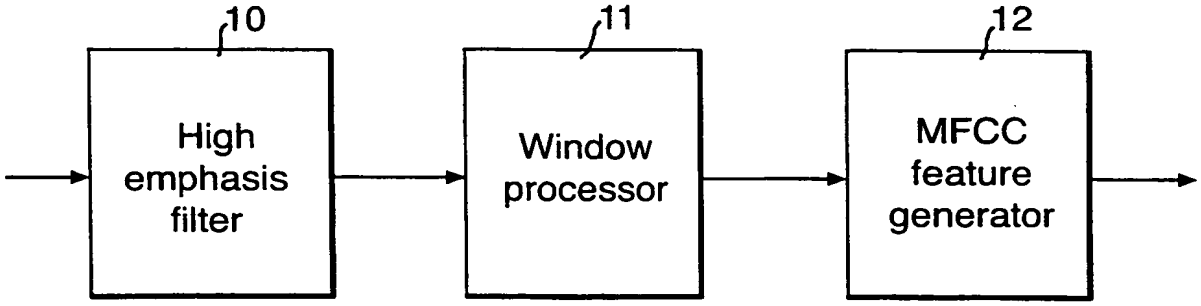

10
High emphasis filter
11
Window processor
12
MFCC feature generator

Fig.5.

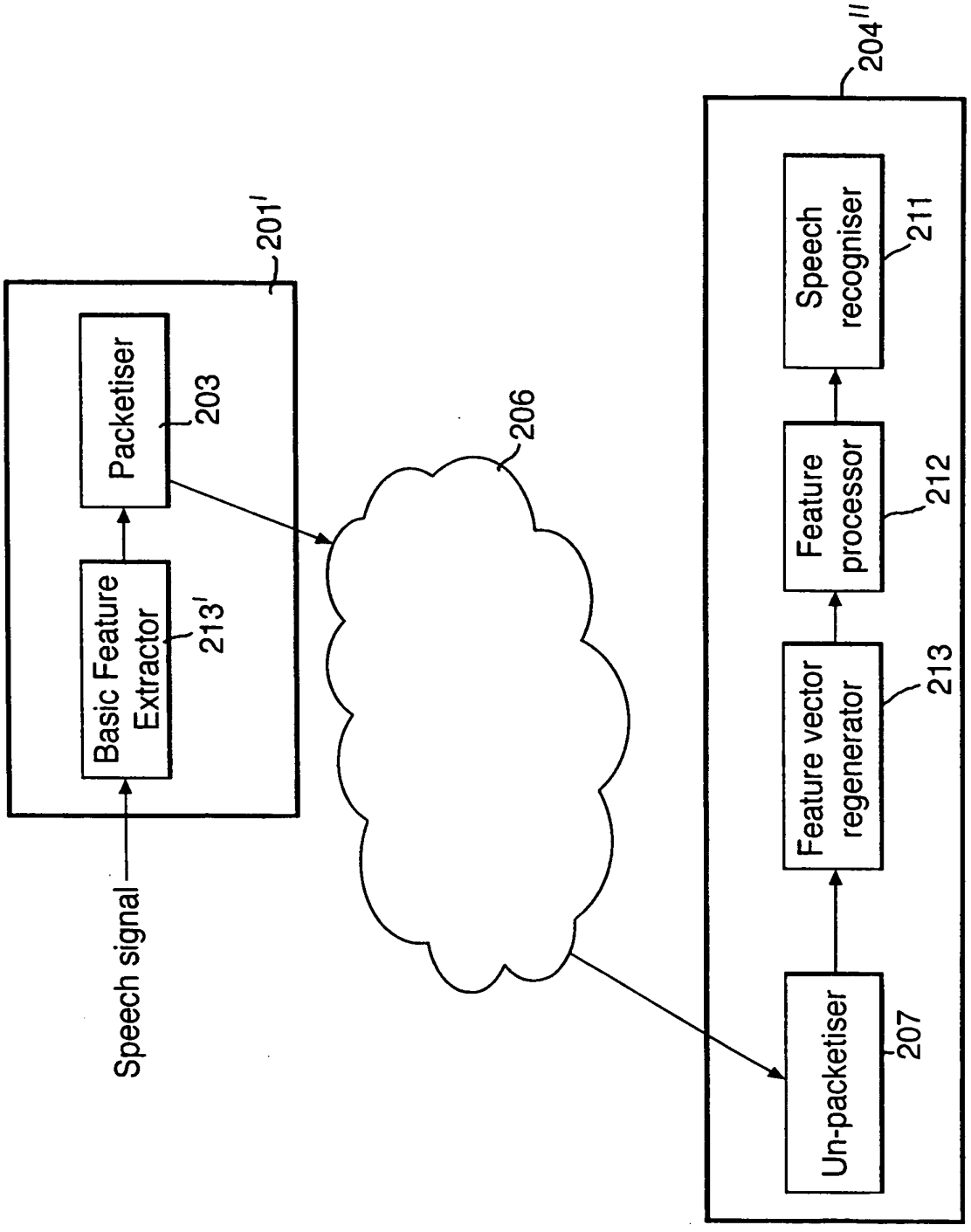
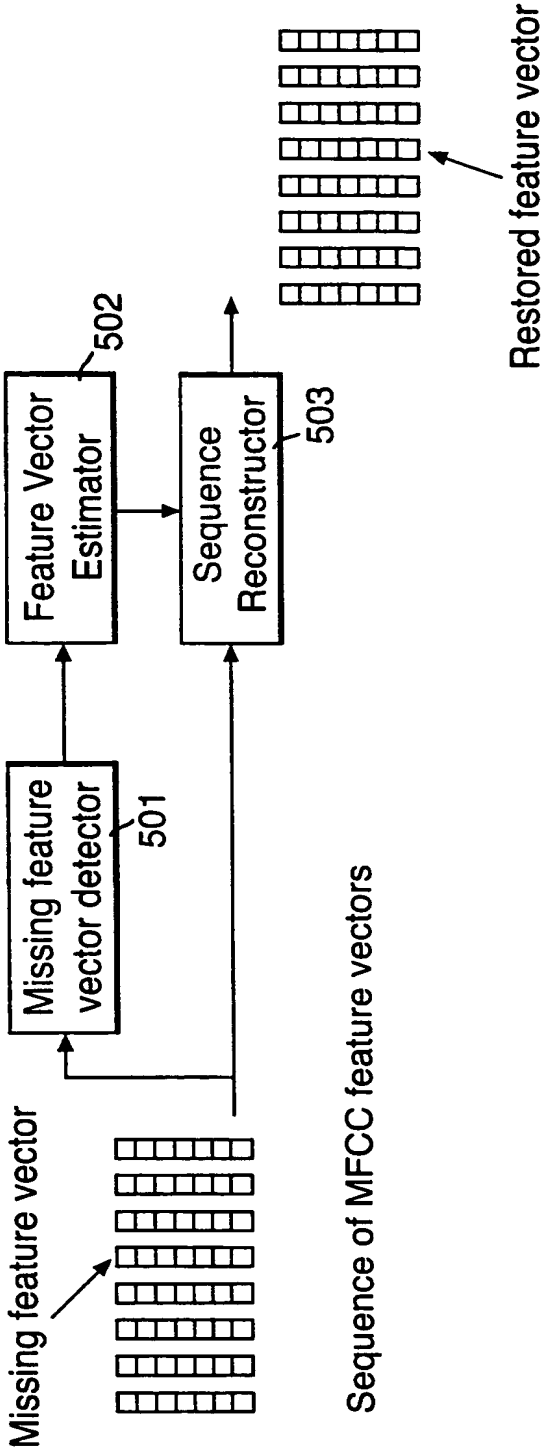

Speech signal
Basic Feature Extractor
213'
Packetiser
203
201'
206
Un-packetiser
207
Feature vector regenerator
213
Feature processor
212
Speech recogniser
211
204''

Fig.6.


Missing feature vector
Missing feature vector detector
501
Feature Vector Estimator
502
Sequence Reconstructor
503
Sequence of MFCC feature vectors
Restored feature vector

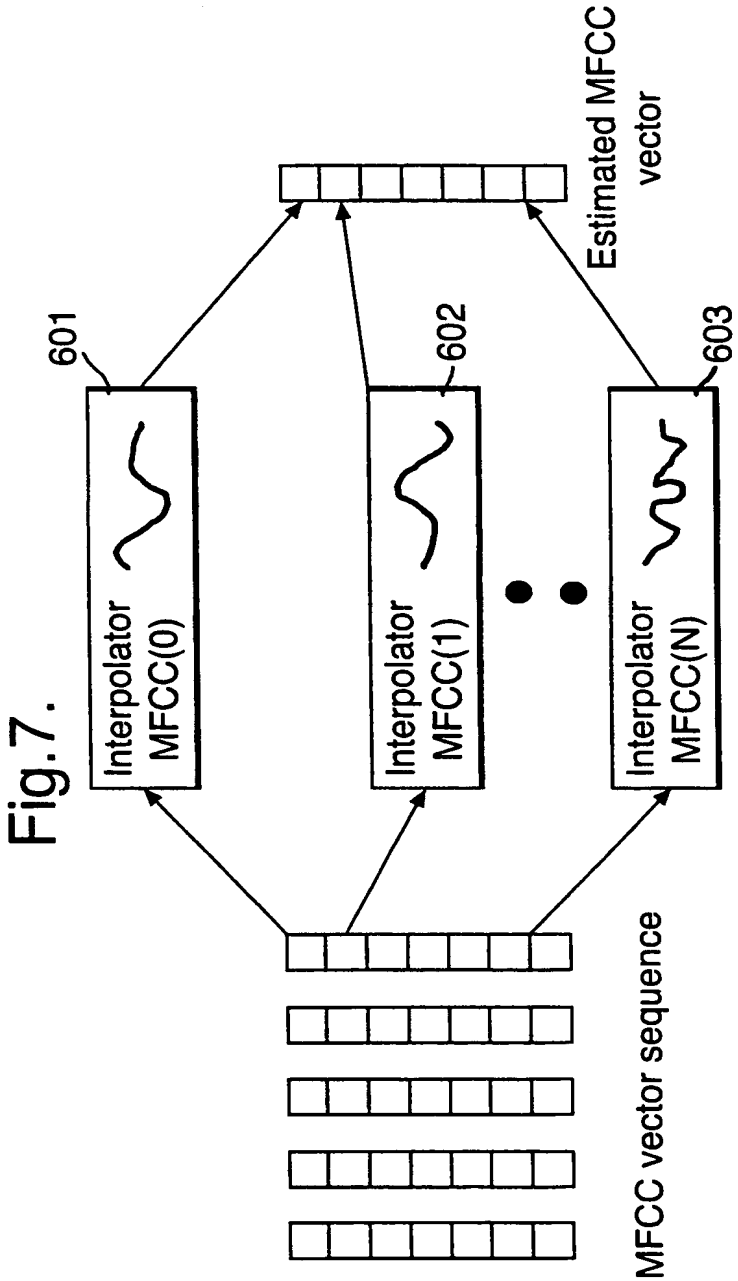
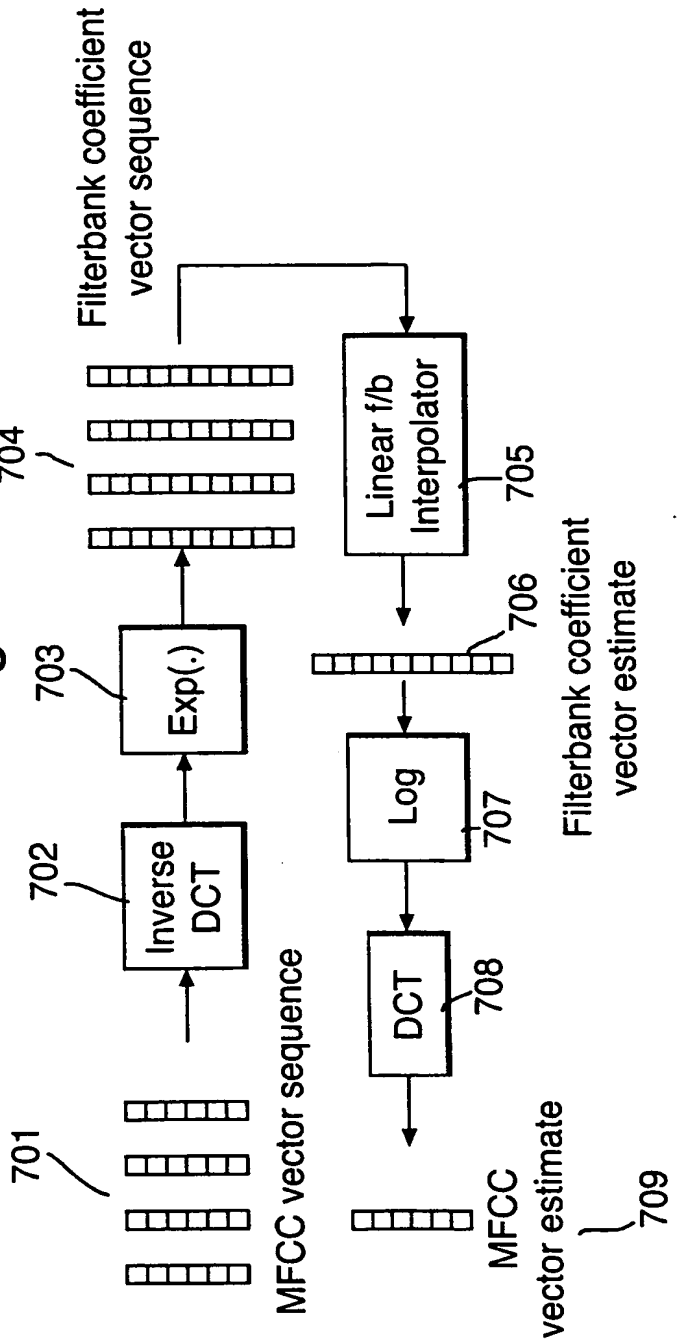

Estimated MFCC vector
601
602
603
Interpolator MFCC(0)
Interpolator MFCC(1)
Interpolator MFCC(N)
MFCC vector sequence

Fig.7.

Fig.8.


Filterbank coefficient vector sequence
704
Linear f/b Interpolator
705
706
Filterbank coefficient vector estimate
703
Exp(.)
Log
707
702
Inverse DCT
DCT
708
701
MFCC vector sequence
MFCC vector estimate
709

14

PATENT COOPERATION TREATY

PCT

REC'D 21 NOV 2001
WIPO PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
A25830 WO		
International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/GB00/04206	02/11/2000	02/11/1999
International Patent Classification (IPC) or national classification and IPC G10L15/26		
Applicant BRITISH TELECOMMUNICATIONS public limited company		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 7 sheets, including this cover sheet.

   ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☒ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☐ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☒ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☒ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
23/04/2001	19.11.2001
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Greiser, N Telephone No. +49 89 2399 7402

Form PCT/IPEA/409 (cover sheet) (January 1994)

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/GB00/04206

I. Basis of the report

1. With regard to the **elements** of the international application *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17))*:
Description, pages:

1-11 as originally filed

Claims, No.:

1-12 as originally filed

Drawings, sheets:

1/7-7/7 as originally filed

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language: , which is:

- ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

- ☐ contained in the international application in written form.
- ☐ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

- ☐ the description, pages:
- ☐ the claims, Nos.:

☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

(Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non-obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 11-12.

because:

☐ the said international application, or the said claims Nos. relate to the following subject matter which does not require an international preliminary examination (*specify*):

☒ the description, claims or drawings (*indicate particular elements below*) or said claims Nos. 11-12 are so unclear that no meaningful opinion could be formed (*specify*):
see separate sheet item VIII

☐ the claims, or said claims Nos. are so inadequately supported by the description that no meaningful opinion could be formed.

☐ no international search report has been established for the said claims Nos. .

2. A meaningful international preliminary examination cannot be carried out due to the failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions:

☐ the written form has not been furnished or does not comply with the standard.

☐ the computer readable form has not been furnished or does not comply with the standard.

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N) Yes: Claims 1-10

	No:	Claims	
Inventive step (IS)	Yes:	Claims	4,8
	No:	Claims	1-3,5-7,9-10
Industrial applicability (IA)	Yes:	Claims	1-10
	No:	Claims	

2. Citations and explanations
see separate sheet

VI. Certain documents cited

1. Certain published documents (Rule 70.10)

and / or

2. Non-written disclosures (Rule 70.9)

see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:
see separate sheet

To Section V:

1. Claim 1 does not meet the requirement of Article 33(3) PCT for the following reason:

Document D1= WO-A-95/17746 discloses a distributed speech recognition system whereby a remote speech recognizer is accessed across a transmission network (see p. 1, par. 1; and p. 6, par. 2; fig. 2). Furthermore, it is known that transmission errors may occur within the transmission network (see p. 10, l. 19-28).
A skilled person would realize that, due to transmission errors, a packet of a received digital signal will occasionally fail to be received or contain errors. To solve the technical problem, the skilled person would consult the relevant prior art. In her search, she would find document D2= US-A-4907277 which discloses a method of reconstructing lost data in a digital voice transmission system.

Document D2 discloses a method for correcting errors due to bits lost in transmission between a transmitter and a receiver; a scheme is disclosed for interpolating or extrapolating bits that were normally received (see col. 1, l. 11-14); as a result , lost samples of a residual baseband signal, an energy component, and/or PARCOR coefficients constituting a speech signal can be reconstructed (see col. 6, l. 16-20).

In detail, D2 discloses a method comprising the steps of
- receiving a sequence of transmitted feature vectors, said feature vectors representing a speech signal (see col. 4, l. 3-8) [PARCOR coefficients are an example for feature vectors];
- detecting the absence of a feature vector in the received sequence (see col. 11, l. 3-8);
- generating an estimated replacement feature vector for the detected absent feature vector (see col. 7, l. 38- col. 9, l. 61); and
- inserting said replacement feature vector into the received feature vector sequence to provide a modified feature vector sequence (see col. 11, l. 38-47).

The skilled person would apply the teaching of D2 upon D1 which would lead in

an obvious manner to the subject-matter of claim 1.
As a result, claim 1 is obvious and, hence, does not involve an inventive step.

2. Claims 2 and 3 do not involve an inventive step (Article 33(3) PCT), respectively.

In particular, D1 discloses a method whereby a generating step comprises estimating a component of the replacement feature vector (see col. 8, l. 46- col. 9, l. 19) (= claim 2).
Moreover, D1 discloses a controller which permanently retains blocks of samples last received. Therefore, the estimating step uses an interpolation coefficient which is updated in response to one or more received feature vectors (see col. 7, l. 65-68) (= claim 3).

3. Claim 4 is new and involves an inventive step. Claim 4 claims an estimating step which converts a received cepstral feature vector into spectral domain for the purpose of interpolation and, thereafter, converts an estimated spectral component back to said cepstral domain. As a result, the subject-matter of claim 4 distinguishes from the prior art and, furthermore, is non-obvious to a skilled person.
Since apparatus claim 8 corresponds to method claim 4, claim 8 is new and inventive as well.

4. Claim 5 and thereof dependent claims 6-7 claim a device for performing speech recognition upon a sequence of parameterised feature vectors. Since apparatus claims 5-7 correspond to method claims 1-3 respectively, the objection of lack of inventive step raised against claims 1-3 applies to claims 5-7 as well.
Consequently, claims 5-7 do not involve an inventive step.

5. Claims 9-10 claim a data carrier when loaded into a computer and carrying out instructions. As far as dependent upon claims 1-3 and 5-7, media claims 9-10 are non-inventive.

To Section VI:

6. Certain published documents (Rule 70.10)

Application No Patent No	Publication date *(day/month/year)*	Filing date *(day/month/year)*	Priority date *(valid claim)* *(day/month/year)*
WO-A-00/30072	May 25, 2000	Nov.12, 1999	Nov. 13, 1998

To Section VIII:

7. Claims 11-12 do not meet the requirement of Rule 6.2(a) PCT since they rely on references to drawings. The used term "as described herein with reference to Figures 5 to 8" is not allowable.

Remarks:

8. The examiner acknowledges the receipt of applicant's reply with letter of Sep. 25, 2001 in response to the first written opinion.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 00/04206

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L15/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, INSPEC, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 00 30072 A (MOTOROLA LTD ;GIBBS JON ALASTAIR (GB); PEARCE DAVID JOHN BENJAMIN) 25 May 2000 (2000-05-25) abstract page 5, line 8 - line 19 page 5, line 26 - line 28 page 11, line 9 - line 15 ---	1,2,4-6, 8-10
Y	WO 95 17746 A (QUALCOMM INC) 29 June 1995 (1995-06-29) abstract page 6, line 6 -page 7, line 3 page 10, line 19 - line 28 ---	1,2,4-6, 8-10
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 January 2001	30/01/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Ramos Sánchez, U

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 00/04206

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	US 4 907 277 A (CALLENS PAUL ET AL) 6 March 1990 (1990-03-06) abstract column 1, line 35 - line 41 column 2, line 10 - line 21	1,2,4-6, 8-10
A	EP 0 872 827 A (AT & T CORP) 21 October 1998 (1998-10-21) column 9, line 27 - line 48; figure 7	1,2,4-6, 8-10
A	PERKINS C ET AL: "A survey of packet loss recovery techniques for streaming audio" IEEE NETWORK, vol. 12, no. 5, September 1998 (1998-09) - October 1998 (1998-10), pages 40-48, XP002133605 IEEE, USA ISSN: 0890-8044 cited in the application page 44 -page 47	1-12
P,X	MILNER B ET AL: "Robust speech recognition over IP networks" IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP 2000), ISTANBUL, TURKEY, 5 - 9 June 2000, pages 1791-1794 vol.3, XP002156639 IEEE, Piscataway, NJ, USA ISBN: 0-7803-6293-4 the whole document	1-12

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 00/04206

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0030072	A	25-05-2000	GB	2343777 A	17-05-2000
			AU	1385600 A	05-06-2000
WO 9517746	A	29-06-1995	AU	692820 B	18-06-1998
			AU	1375395 A	10-07-1995
			BR	9408413 A	05-08-1997
			CA	2179759 A	29-06-1995
			CN	1138386 A	18-12-1996
			EP	0736211 A	09-10-1996
			FI	962572 A	20-08-1996
			JP	9507105 T	15-07-1997
			US	5956683 A	21-09-1999
			ZA	9408426 A	30-06-1995
US 4907277	A	06-03-1990	EP	0139803 A	08-05-1985
			CA	1245780 A	29-11-1988
			DE	3374109 D	19-11-1987
			JP	1629175 C	20-12-1991
			JP	2055880 B	28-11-1990
			JP	60098600 A	01-06-1985
EP 0872827	A	21-10-1998	US	6078886 A	20-06-2000
			CA	2228917 A	14-10-1998
			JP	10333693 A	18-12-1998

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Commissioner
US Department of Commerce
United States Patent and Trademark
Office, PCT
2011 South Clark Place Room
CP2/5C24
Arlington, VA 22202
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 03 August 2001 (03.08.01)	
International application No. PCT/GB00/04206	**Applicant's or agent's file reference** A25830 WO
International filing date (day/month/year) 02 November 2000 (02.11.00)	**Priority date** (day/month/year) 02 November 1999 (02.11.99)
Applicant MILNER, Benjamin, Peter	

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International Preliminary Examining Authority on:

23 April 2001 (23.04.01)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO **34, chemin des Colombettes** **1211 Geneva 20, Switzerland**	Authorized officer Zakaria EL KHODARY
Facsimile No.: (41-22) 740.14.35	Telephone No.: (41-22) 338.83.38